FOODBORNE PATHOGENS AND DISEASE
Volume 6, Number 3, 2009

DOI: 10.1089/fpd.2008.0185

Genotyping *Campylobacter jejuni* by Comparative Genome Indexing: An Evaluation with Pulsed-Field Gel Electrophoresis and *flaA* SVR Sequencing

Lauren G. Pittenger,[1] Mark D. Englen,[1] Craig T. Parker,[2]
Jonathan G. Frye,[1] Beatriz Quiñones,[2] Sharon T. Horn,[2] Insook Son,[1]
Paula J. Fedorka-Cray,[1] and Mark A. Harrison[3]

Abstract

Comparative genome indexing (CGI) using whole-genome DNA microarrays was evaluated as a means of genotyping *Campylobacter jejuni* relative to two standard methods, pulsed-field gel electrophoresis (PFGE) and *flaA* short variable region sequencing (*flaA* SVR typing). Thirty-six geographically diverse *C. jejuni* isolates were selected from a collection of cattle and chicken isolates. The BioNumerics® software program was used for cluster analysis of the data from all 36 isolates for each of the three typing methods. Comparative genome indexing assigned a unique type to each isolate while PFGE and *flaA* SVR distinguished 29 and 35 different types, respectively. The four common types identified by PFGE were also closely related by CGI, and the overall similarity of the CGI results to those for PFGE indicates the value of CGI as a more informative alternative to PFGE. While *flaA* SVR was very discriminative, the isolates were all highly similar (>78%) resulting in finer distinctions between isolates and fewer genotypic relations to CGI or PFGE. *Campylobacter jejuni* is one of the most common causative agents of bacterial gastroenteritis in the world. The development of CGI as a molecular typing tool for *C. jejuni* offers a highly effective and informative means of further understanding the epidemiology of this ubiquitous pathogen.

Introduction

Campylobacter is a common zoonotic pathogen. Food animals, particularly poultry, comprise the major reservoir for this organism (Moore *et al.*, 2005). In humans, *Campylobacter* infections are most often sporadic in nature, although outbreaks do occur. In the United States, *Campylobacter* infections causing diarrheal disease have been reported to occur 1.5 to 5.8 times more frequently than those due to *Salmonella*, *Shigella*, or *Escherichia coli* O157:H7 (Friedman *et al.*, 2000). It is estimated that in the United States there are approximately 2.5 million illnesses and over 120 deaths due to *Campylobacter* spp. every year (Mead *et al.*, 1999). Among *Campylobacter* spp., *C. jejuni* and to a lesser extent *C. coli* are most commonly isolated from human infections (Skirrow, 1994).

Effective discriminatory typing methods are essential for source tracking and epidemiological studies of major foodborne pathogens such as *Campylobacter*. Methods used for genotyping *Campylobacter* include amplified fragment length polymorphism, flagellin typing (*fla* typing), pulsed-field gel electrophoresis (PFGE),

[1]Bacterial Epidemiology and Antimicrobial Resistance Research Unit, Agricultural Research Service, U.S. Department of Agriculture, Athens, Georgia.
[2]Produce Safety and Microbiology Research Unit, Agricultural Research Service, U.S. Department of Agriculture, Albany, California.
[3]Department of Food Science and Technology, University of Georgia, Athens, Georgia.

ribotyping, and random amplified polymorphic DNA analysis (Wassenaar and Newell, 2000). PFGE and *fla* typing are two of the most commonly used and well-accepted methods for genotyping *C. jejuni* (Patton *et al.*, 1991). Among typing methods that are based on the band patterns of DNA fragments, PFGE is one of the most discriminatory (Gibson *et al.*, 1997). However, PFGE has several disadvantages, including the high cost of reagents and the large amount of time required to prepare the samples. In addition, appropriate restriction enzymes must be identified empirically because no single enzyme adequately digests the DNA of all bacterial strains. Further, as is the case with all DNA fragment–based methods, interpreting and standardizing the results is sometimes difficult and subject to error (Tenover *et al.*, 1995; Singer *et al.*, 2004).

Analysis of the DNA sequence variability of the short variable region (SVR) of the *flaA* flagellin gene has proven to be a simple and useful variation of the *fla* typing method for *Campylobacter* allowing relatively high sample throughput at reasonable cost (Meinersmann *et al.*, 1997, 2005). This technique uses PCR primers that hybridize to conserved sequences flanking the *flaA* SVR to amplify a 267-bp region for sequencing. Sequence-based *flaA* typing avoids difficulties inherent in methods that rely on restriction fragment length polymorphisms of the flagellin genes. In particular, the gel patterns produced by restriction digests of *flaA* PCR products are subject to variability and the information relating to strain differences that can be obtained is limited (Wassenaar and Newell, 2000). Since *flaA* SVR is limited to analysis of variations in a single gene, long-term time–location trends cannot be examined, but this method can be very useful for discriminating more closely related *Campylobacter* isolates (Hiett *et al.*, 2007).

Over the last decade the sequences of a large number of bacterial genomes have been completed, including several species of *Campylobacter* (http://www.ncbi.nlm.nih.gov/genomes/lproks.cgi). The complementary development of genome-scale hybridization technology employing DNA microarrays has enabled detailed analyses of the genetic composition of a wide range of microorganisms, including *Campylobacter* (Brown and Botstein, 1999; Ehrenreich, 2006). Comparative genome indexing (CGI) using whole-genome DNA microarrays allows discrimination of bacterial isolates based upon comparisons of the gene content of each strain to that of a reference strain or strains on the microarray (Anjum *et al.*, 2003; Call *et al.*, 2003). This method has been used to investigate the global genomic diversity of *C. jejuni* strains and offers considerable potential as a genotyping tool since it provides more strain-specific information than other typing techniques currently in use (Leonard *et al.*, 2003; Pearson *et al.*, 2003; Taboada *et al.*, 2004; Parker *et al.*, 2006; Quiñones *et al.*, 2008). However, CGI is not without limitations. Point mutations, small deletions, gene rearrangements, and intergenic regions containing promoter elements and nontranslated RNAs are typically not present on DNA microarrays and are thus excluded from the analysis. As with PFGE, CGI is more costly and technically demanding than other genotyping methods, and the analysis is dependent on the genomes incorporated into the microarray design. Despite this, CGI has emerged as an important new means of characterizing microbial populations, providing insights into microbial evolution and genetic diversity (Ochman and Santos, 2005; Garaizar *et al.*, 2006; Kostrzynska and Bachand, 2006).

Thus far no typing method has been determined to be the definitive choice for *Campylobacter*. Moreover, typing methods based on very different strategies may be expected to yield differing results and conclusions. This study was designed to compare two of the most commonly used methods for genotyping *C. jejuni*, PFGE and *flaA* SVR, to CGI using whole-genome DNA microarrays, and to determine whether CGI is as discriminatory as the conventional methods. A geographically diverse group of 36 *C. jejuni* isolates from chickens and beef cattle representing five main regions of the United States were selected and typed using CGI, PFGE, and *flaA* SVR.

Materials and Methods

Bacterial isolates and culture conditions

Eighteen *C. jejuni* isolates from chicken carcass rinses and 18 *C. jejuni* isolates from beef cattle feces were selected from the *Campylobacter* collection at the U.S. Department of

Agriculture-Agricultural Research Service Russell Research Center in Athens, Georgia. Isolates were chosen for this study from three different years (1999, 2001, and 2002) and from all five major geographic regions of the continental United States to provide isolate diversity. These regions were divided as follows: Midwest (Iowa, Illinois, Kansas, Minnesota, Missouri, North Dakota, Nebraska, South Dakota, and Wisconsin), Northeast (Connecticut, Delaware, Indiana, Massachusetts, Maryland, Maine, Michigan, New Hampshire, New York, New Jersey, Ohio, Pennsylvania, Rhode Island, and Vermont), Southeast (Alabama, Florida, Georgia, Kentucky, North Carolina, Puerto Rico, South Carolina, Tennessee, Virginia, and West Virginia), Southwest (Arkansas, Louisiana, Mississippi, Oklahoma, and Texas), and West (Arizona, California, Colorado, Idaho, Montana, New Mexico, Nevada, Oregon, Utah, Washington, and Wyoming). The isolates were not epidemiologically linked and were geographically independent. *Campylobacter* isolates were routinely stored as frozen stocks at −80°C in Mueller-Hinton broth supplemented with 10% glycerol. The isolates were recovered from frozen stocks on Campy-Cefex agar plates (Stern *et al.*, 1992) and incubated under microaerobic conditions (5% O_2, 10% CO_2, and 85% N_2) in zip-top bags for 48 hours at 42°C. Species confirmation of the isolates was determined using the *Campylobacter* BAX® PCR (DuPont™ Qualicon, Wilmington, DE) as previously described (Englen and Fedorka-Cray, 2002).

Campylobacter jejuni *DNA microarray construction*

The DNA microarrays used in this study were constructed from the open reading frames (ORFs) of *C. jejuni* strains NCTC11168 (Parkhill *et al.*, 2000) and RM1221 (Fouts *et al.*, 2005). For NCTC11168, a total of 1530 individual ORFs were amplified using the Sigma Genosys (The Woodlands, TX) *C. jejuni* ORFmer primer set, and for RM1221, a total of 227 individual ORFs were amplified with primers from Operon Technologies (Alameda, CA) that were designed with ArrayDesigner (Version 2.0, Premier Biosoft, Palo Alto, CA) as in previous studies (Parker *et al.*, 2006; Quiñones *et al.*, 2008). The PCR products were purified on a Qiagen 8000 robot by using a Qiaquick 96-well Biorobot kit (Qiagen, Valencia, CA) and spotted in duplicate onto Ultra-GAPS glass slides (Corning Inc., Corning, NY) using an OmniGrid Accent (GeneMachines, Ann Arbor, MI), as described previously (Parker *et al.*, 2006; Quiñones *et al.*, 2008). Immediately after printing, the microarrays were UV crosslinked and blocked (Parker *et al.*, 2006; Quiñones *et al.*, 2008).

Isolation and fluorescent labeling of genomic DNA

Genomic *C. jejuni* DNA was isolated using the PUREGENE Yeast & Gram-Positive Bacteria kit (Gentra Systems, Minneapolis, MN) or the Qiagen DNeasy™ kit, according to the manufacturer's directions. For each microarray hybridization reaction, DNA from the reference strains (*C. jejuni* strain NCTC11168 and *C. jejuni* strain RM1221) and a test isolate were fluorescently labeled with indodicarbocyanine (Cy5) and indocarbocyanine (Cy3), respectively. Approximately 2 μg of DNA was mixed with 5 μL 10×NEBlot labeling buffer containing random sequence octamer oligonucleotides (New England Biolabs, Beverly, MA) and water to a final volume of 41 μL. This mixture was heated to 95°C for 5 minutes and then cooled for 5 minutes at 4°C. Following this, the remainder of the labeling reaction components were added: 5 μL of 10×dNTP labeling mix (1.2 mM each dATP, dGTP, dCTP plus 0.5 mM dTTP in 10 mM Tris pH 8.0 and 1 mM EDTA), 3 μL of Cy3 dUTP or Cy5 dUTP (Amersham Biosciences, Piscataway, NJ) and 1 μL of Klenow fragment (New England Biolabs). The labeling reactions were incubated overnight at 37°C. Labeled DNA was purified from unincorporated label using Qiaquick PCR Cleanup kits (Qiagen) and dried by vacuum.

Microarray hybridization and data analysis

Labeled reference and test DNA were combined in a 45 μL Pronto! cDNA hybridization solution (Corning Life Sciences, Corning, NY) and heated to 95°C for 5 minutes and 15 μL of the hybridization mixture was put onto a slide, sealed with a coverslip and incubated at 42°C for 18 hours. Following hybridization, the slides were washed and dried before scanning,

as reported previously (Parker *et al.*, 2006; Quiñones *et al.*, 2008). At least two hybridization reactions were performed for each test isolate and were quantified as described below.

DNA microarrays were scanned using an Axon GenePix 4000B microarray laser scanner (Molecular Devices, Sunnyvale, CA) at 532 nm (Cy3) and 635 nm (Cy5) excitation wavelengths. Features and local background intensities were measured and quantified using GenePix 4.0 software (Molecular Devices). Spots were excluded from further analysis if they contained anomalous spot morphology or were within regions of high-fluorescent background (Quiñones *et al.*, 2008). The data were filtered so that spots with a reference signal lower than the background plus two standard deviations of the background were discarded. To compensate for unequal dye incorporation, data normalization was performed as described previously (Parker *et al.*, 2006; Quiñones *et al.*, 2008). The normalized data were converted to Cy3/Cy5 ratio data for all isolates and analyzed further with GeneSpring microarray analysis software (Version 6.2, Agilent Technologies, Santa Clara, CA) using average-linkage hierarchical clustering with standard correlation and bootstrapping (Parker *et al.*, 2006; Quiñones *et al.*, 2008). The CGI analysis defined the status of a gene as present when the normalized Cy3/Cy5 ratios (test isolate/reference strain) were $\geq$0.5 and variable (i.e., absent or present but with sufficient sequence divergence to prevent hybridization) when the normalized Cy3/Cy5 ratios were $<$0.5. The normalized intensity ratios were converted into binary form using an average of the four data points for each gene. A value of 1 was assigned to genes that were present, and a value of 0 was assigned to those found to be variable. These binary values were then imported into the BioNumerics® software program (Version 3.5, Applied Maths, Austin, TX) using a script designed specifically for this purpose. The Pearson similarity coefficient and the unweighted pair group method using arithmetic averages (UPGMA) were used for cluster analysis.

PFGE

The 24-hour standardized *Campylobacter* PulseNet protocol of the U.S. Centers for Disease Control and Prevention (Ribot *et al.*, 2001) was used with one modification: the lysis step was extended from 15 minutes to 2 hours. All *Campylobacter* isolates were digested with *Sma*I restriction enzyme; *Salmonella* ser. Braenderup H9812 DNA digested with *Xba*I was used as the molecular size standard.

The *Campylobacter* PFGE patterns were analyzed with BioNumerics® using an optimization setting of 1.0% and band position tolerance of 0.5%. Suspected double bands were resolved by examining the plotted densitometric profiles of the band patterns. Cluster analysis was performed using the Dice coefficient and UPGMA. In addition, all PFGE band patterns were carefully determined to be correctly grouped by visual inspection.

flaA *SVR typing*

The method for *flaA* SVR typing described by Meinersmann *et al*. (1997) was used with some modification. Genomic DNA was isolated using the PUREGENE DNA isolation kit (Gentra Systems). The SVR region of the *C. jejuni flaA* gene was amplified using the FLA4F and FLAA623RU primer pair with the 3′ primer outside of the SVR to ensure that only the *flaA* gene was amplified. For the sequencing reactions, the FLA242FU and FLA623RU primers were used. The sequencing reactions were performed on an ABI 9700 thermal cycler (Applied Biosystems, Foster City, CA) using the ABI PRISM BigDye terminator cycle sequencing kit (Version 1.1) according to the manufacturer's standard protocols. The DNA fragments were separated and analyzed on an ABI 3100 Genetic Analyzer with ABI PRISM Genetic Analyzer data collection software (Applied Biosystems). Sequences were further analyzed and contigs were aligned using Vector NTI Advance software (Version 9.0; Invitrogen, Carlsbad, CA). Sequence data were then imported into BioNumerics and analyzed using pairwise comparisons.

Results

PFGE

PFGE differentiated 25 unique and 4 common types among the 36 *C. jejuni* isolates

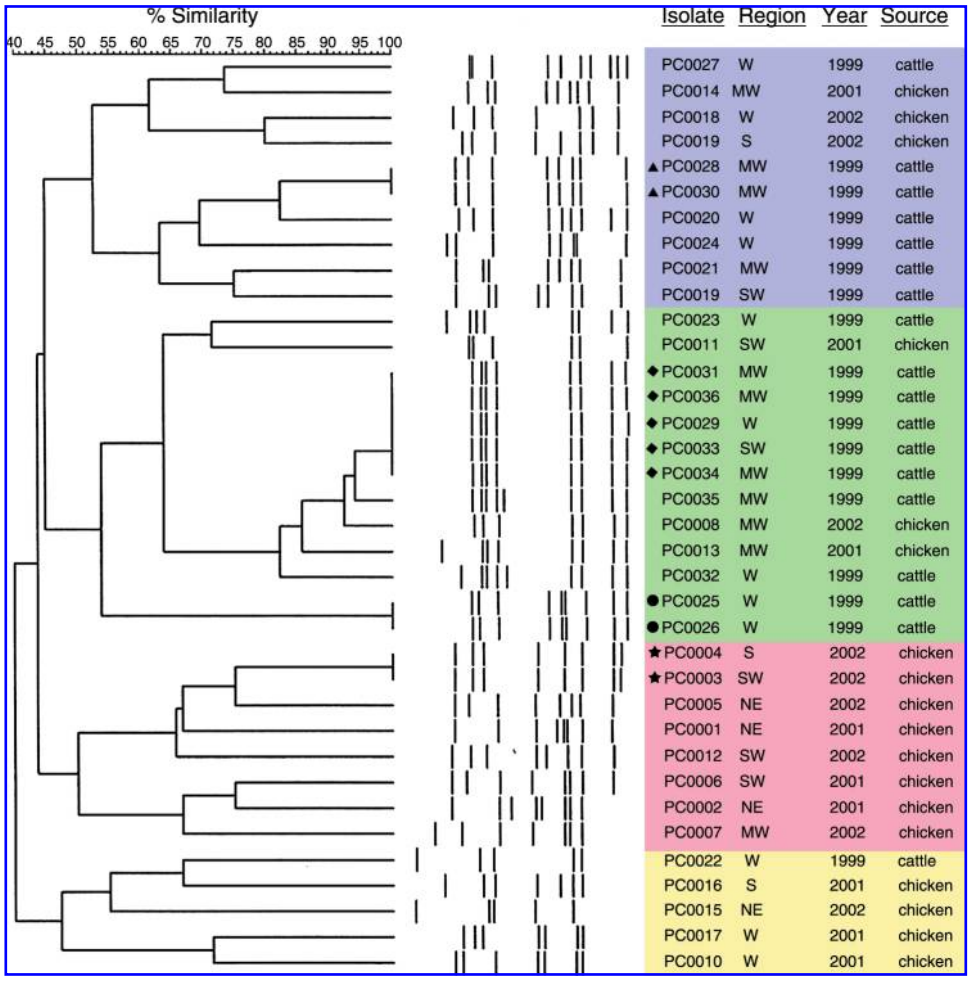

FIG. 1. The genetic relatedness of cattle and chicken *C. jejuni* strains determined by pulsed-field gel electrophoresis (PFGE). The dendrogram was generated in Bionumerics from cluster analysis using UPGMA and the Dice similarity coefficient. Shaded blocks indicate major PFGE clusters. Symbols (▲◆●★) designate common PFGE types. Regions: MW, Midwest; NE, Northeast; S, South; SW, Southwest; W, West.

investigated. Four major clusters were identified, highlighted by colored shading in Fig. 1. The largest of these clusters (green shading, Fig. 1) included a group with five common-type cattle isolates (filled diamonds, Fig. 1) and four additional isolates, two from chickens (PC0008 and PC0013) and two from cattle (PC0032 and PC0035). Two pairs of outlying isolates were also part of this cluster: chicken isolate PC0011 and cattle isolate PC0023, and two cattle isolates with common types (filled circles, Fig. 1). Isolates grouped mainly by host source, illustrated by the group of six cattle isolates (PC0019–PC0021, PC0024, PC0028, and PC0030) in the second largest major cluster (blue shading, Fig. 1), and the major cluster composed entirely of *C. jejuni* isolates from chickens (red shading, Fig. 1); the latter cluster included the only PFGE common type composed of chicken isolates (filled stars, Fig. 1). The smallest of the main PFGE clusters was also composed primarily of chicken isolates (yellow shading, Fig. 1). In contrast to isolate grouping by host source, little distinct clustering by geographic region was observed.

CGI

Comparative genome indexing differentiated all of the 36 *C. jejuni* isolates as unique types; the genotypic relationships of these CGI types relative to those determined by PFGE are

illustrated by colored shading in Fig. 2. A comparison with Fig. 1 shows that the main isolate clusters determined using PFGE were largely preserved by CGI although changes in specific genotypic relationships among isolates were apparent. For example, the 10-isolate major PFGE cluster (blue shading, Fig. 1) consisted of two smaller, four-isolate (cattle isolate PC0027 plus chicken isolates PC0009, PC0014, and PC0018) and six-isolate (cattle isolates PC0019-PC0021, PC0024, PC0028, and PC0030) clusters. With CGI, 8 of these 10 isolates remained grouped together but their genotypic relationships were quite different (blue shading, Fig. 2); note that cattle isolates PC0028 and PC0030 had a common PFGE type but unique types by CGI. Also, chicken isolate PC0009 and cattle isolate PC0020 were split out of this PFGE group by CGI (blue shading, Figs. 1 and 2). Similarly, the 13-isolate PFGE cluster (green shading, Fig. 1), containing two common cattle types (filled diamonds and circles, Fig. 1), became more complex by CGI. For instance, the five cattle isolates with a common PFGE type (filled diamonds, Figs. 1 and 2) remained closely related by CGI but each had a unique type. Also, the common PFGE type of cattle isolates PC0025 and PC0026 (filled circles, Figs. 1 and 2) was differentiated by CGI and related more closely to chicken isolates PC0009, PC0011, and PC0014 (blue, green, and yellow shading, respectively, Fig. 2). Chicken isolate PC0008 was separated by CGI from its small PFGE cluster, a result similar to that for cattle isolate PC0023.

The major PFGE cluster consisting only of chicken isolates (PC0001–PC0007 and PC0012) remained intact by CGI (red shading, Figs. 1 and 2) though the genotypic relationships of the isolates within these clusters were somewhat different; isolates PC0003 and PC0004 that had a common PFGE type were differentiated by CGI. Chicken isolates PC0010 and PC0017 maintained their relative relationship by PFGE and CGI; however, chicken isolates PC0015–PC0016 and cattle isolate PC0022 separated from this group by CGI (yellow shading, Figs. 1 and 2). As with PFGE, both cattle and chicken *C. jejuni* isolates tended to cluster by host source but not by geographic region with CGI (Fig. 2).

The genomic regions of the *C. jejuni* isolates with greatest variability and that therefore primarily differentiated the isolates by CGI are shown in Table 1. A total of 470 variable genes (276 from *C. jejuni* NCTC11168 and 194 from *C. jejuni* RM1221) distributed within 11 main functional categories were identified. Most of the variable genes identified from NCTC11168 (47.8%, 132/276) covered a broad range of functional categories and were therefore grouped as mixed in Table 1 for the sake of simplicity. Cell envelope components, hypothetical proteins, and ORFs with unknown functional properties also represented significant (13–15%, 35–40 of 276) portions of the total NCTC11168 variable genes. In contrast, the majority (62%, 120/194) of variable genes from *C. jejuni* RM1221 consisted of genomic islands (the Mu-like prophage CJIE1 and integrated elements CJIE2 and CJIE4). However, cell envelope components accounted for a percentage of the variable genes from RM1221 similar to that for NCTC11168 (12%, 23/194 and 13%, 35/276, respectively). Within these groups of variable genes, all or a portion of 17 of the 18 intraspecies hypervariable regions published by Parker *et al.* (2006) were identified; region 18 (CJE0944–CJE0947) was absent in our analysis. Portions of two additional recently reported (Rodin *et al.*, 2008) variable regions (regions 19 and 21) were also among the variable genes in our data.

flaA *SVR typing*

The *flaA* SVR results produced a dendrogram with 35 different types (cattle isolates PC0023 and PC0030 had a common *flaA* SVR type) though all isolates had a greater similarity (>78%) than with PFGE or CGI. The genotypic relationships of the *flaA* SVR types relative to those determined by PFGE are illustrated by colored shading in Fig. 3. The 36 *C. jejuni* isolates were distributed among two major clusters, separated by a smaller group containing a chicken isolate (PC0014) and two cattle isolates (PC0019 and PC0021). Note that this latter group of three isolates was largely preserved by all three typing methods (blue shading, Figs. 1–3).

The major nine-member *flaA* SVR cluster consisting of seven cattle and two chicken isolates (green shading, Fig. 3) was also an integral component of a corresponding PFGE cluster

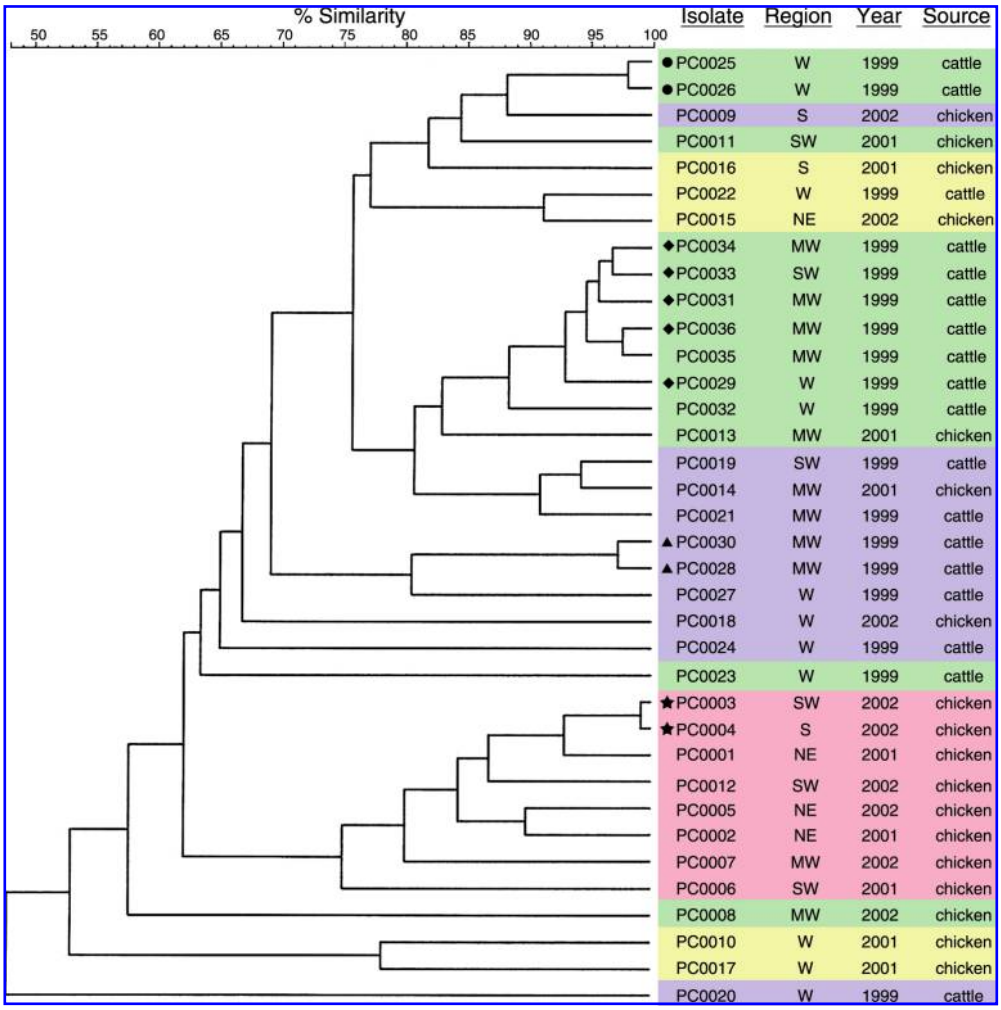

FIG. 2. The genetic relatedness of cattle and chicken *C. jejuni* strains by comparative genome indexing (CGI). The dendrogram was generated in Bionumerics from cluster analysis using UPGMA and the Pearson similarity coefficient. Shaded blocks indicate isolate relationships relative to the major pulsed-field gel electrophoresis (PFGE) clusters. Symbols (▲◆●★) designate common PFGE types. Regions: MW, Midwest; NE, Northeast; S, South; SW, Southwest; W, West.

although their structures differed (green shading, Figs. 1 and 3); the five cattle isolates with a common PFGE type were discriminated by *flaA* SVR (filled diamonds, Figs. 1 and 3). This major cluster (minus chicken isolate PC0008) was also a feature of the CGI dendrogram (green shading, Fig. 2) and the five PFGE common-type cattle isolates were also discriminated as unique types by CGI (filled diamonds, Figs. 1 and 2).

The largest *flaA* SVR major cluster was a complex rearrangement of the genotypic relationships determined by PFGE and CGI (colored shading, Figs. 1–3). In several instances isolates had closely related genotypes by *flaA* SVR (e.g., chicken isolates PC0003–PC0004 plus cattle isolates PC0020 and PC0024; chicken isolates PC0002, PC0011, and PC0016; chicken isolates PC0006, PC0010, and PC0017) but were quite disparate by PFGE and CGI. The one common *flaA* SVR type consisted of cattle isolates PC0023 and PC0030, both of which were more distantly related by PFGE and CGI (Figs. 1–3). However, the PFGE common type consisting of chicken isolates PC0003 and PC0004 remained closely related by both *flaA* SVR and CGI (filled diamonds, Figs. 1–3) as did the PFGE common type of cattle isolates PC0028 and PC0030 (filled triangles, Figs. 1–3). Cattle

Table 1. Variable Genomic Regions in *Campylobacter jejuni*

Main functional category	*No. of genes*	*Gene map regions*[a]
Acyl carrier proteins	11	Cj1299–Cj1300, Cj1302–Cj1310
Cation ABC transport	9	Cj0136–Cj0144
Cell envelope	35	Cj0057–Cj0058, Cj0176–Cj0178, Cj0183, Cj0420–Cj0421, Cj0423–Cj0424, Cj0628–Cj0629, Cj0817–Cj0819, Cj0859–Cj0860, Cj0986–Cj0988, Cj1135–Cj1144, Cj1160, Cj1443–Cj1448, Cj1677, Cj1679
Energy metabolism	9	Cj0264–Cj0265, Cj0533–Cj0538, Cj0570
Hypothetical proteins	40	Cj0055–Cj0056, Cj0170–Cj0172, Cj0417–Cj0419, Cj0566, Cj0568–Cj0569, Cj0628–Cj632, Cj0735–Cj0742, Cj0747, Cj0752, Cj0814–Cj0816, Cj0961–Cj0965, Cj0967–Cj0973, Cj1429, Cj1433
Mixed	132	Cj0259–Cj0263, Cj0291, Cj0293–Cj0306, Cj0480–Cj0489, Cj0493–Cj0494, Cj0497–Cj0501, Cj0672, Cj0676–Cj0679, Cj0753, Cj0755, Cj0757–Cj0765, Cj1040–Cj1059, Cj1290–Cj1296, Cj1313–Cj1441, Cj1549–Cj1553, Cj1555–Cj1556, Cj1558, Cj1560–Cj1563, Cj1720–Cj1729, Cj1731
Transport and binding proteins	11	Cj0173–Cj0175, Cj0179–Cj0182, Cj0613–Cj0616
Unknown	29	Cj0422, Cj1158–Cj1159, Cj1415–Cj1428, Cj1430–Cj1432, Cj1434–Cj1435
CJIE1 (CMLP1)	42	CJE0215–CJE0216, CJE0220–CJE0222, CJE0226–CJE0234, CJE0236–CJE0237, CJE0241, CJE0243–CJE0252, CJE0254, CJE0256– CJE0259, CJE0261–CJE0263, CJE0265–CJE0266, CJE0268–CJE0270, CJE0272–CJE0273
CJIE2	37	CJE0544, CJE0546, CJE0548–CJE551, CJE0553–CJE0556, CJE0559, CJE0562, CJE0565–CJE0569, CJE0571–CJE0576, CJE0578, CJE0582–CJE0585, CJE0588, CJE0591–CJE0594, CJE0597–CJE0598, CJE0600–CJE0601
CJIE4	41	CJE1418, CJE1421–CJE1423, CJE1425–CJE1427, CJE1430, CJE1433–CJE1436, CJE1439–CJE1442, CJE1444–CJE1445, CJE1447–CJE1453, CJE1455, CJE1457–CJE1462, CJE1465–CJE1469, CJE1471–CJE1474
Cell envelope	23	CJE0310, CJE0314, CJE0387–CJE0389, CJE0466, CJE0472, CJE1278–CJE1281, CJE1284, CJE1602–CJE1604, CJE1606–CJE1607, CJE1611–CJE1616
Hypothetical proteins	37	CJE0045, CJE0139, CJE0664, CJE0673, CJE1093–CJE1094, CJE1100, CJE1105–CJE1107, CJE1111, CJE1113–CJE1114, CJE1117, CJE1119, CJE1122–CJE1123, CJE1125–CJE1126, CJE1128, CJE1131–CJE1132, CJE1135–CJE1138, CJE1141–CJE1143, CJE1146–CJE1148, CJE1150–CJE1154
Mixed	14	CJE0031, CJE0140, CJE0196, CJE0670, CJE0731–CJE0732, CJE1498, CJE1500–CJE1502, CJE1720, CJE1725, CJE1727–CJE1728

[a]As reported by Parkhill *et al.* (2000) for *C. jejuni* NCTC11168 (Cj-) and by Fouts *et al.* (2005) for *C. jejuni* RM1221 (CJE-).

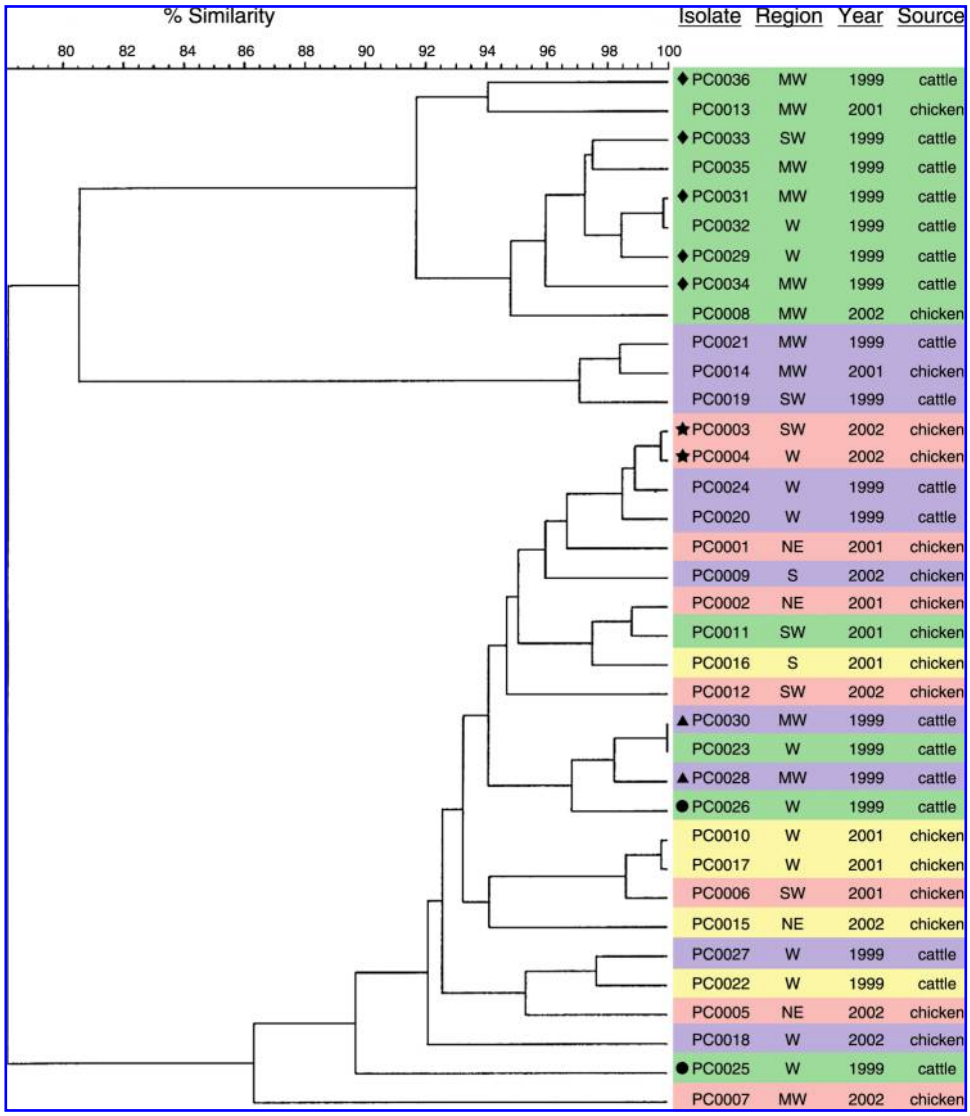

FIG. 3. The genetic relatedness of cattle and chicken *C. jejuni* strains using *flaA* SVR typing. The dendrogram was generated in Bionumerics from cluster analysis using UPGMA and pairwise comparisons. Shaded blocks indicate isolate relationships relative to the major pulsed-field gel electrophoresis (PFGE) clusters. Symbols (▲◆●★) designate common PFGE types. Regions: MW, Midwest; NE, Northeast; S, South; SW, Southwest; W, West.

isolates PC0025–PC0026 had a common PFGE type and were also closely related by CGI, but were placed in separate subgroups by *flaA* SVR (filled circles, Figs. 1–3). Isolates did not cluster by geographic region with *flaA* SVR, and the marked clustering by host source observed with PFGE and CGI was limited to only some of the cattle isolates (Fig. 3).

Discussion

A number of different methods have been used for typing *Campylobacter* (Wassenaar and Newell, 2000), but the Centers for Disease Control and Prevention has adopted PFGE for its national PulseNet program (http://www.cdc.gov/pulsenet) and this method has become the de facto standard for typing *Campylobacter* in the United States. Among the 36 *C. jejuni* isolates in our study, PFGE identified 25 unique types while 11 isolates (31%) belonged to four common types. The majority of common-type isolates were from cattle (9/11, 82%). This reflects the observation that the majority of isolates clustered by host source using PFGE, but also the limitations of PFGE for discriminating

closely related isolates. PFGE depends solely on differences in the genomic locations of restriction enzyme sites; mutations, gene inversions, or minor insertions and deletions not affecting these restriction sites will not be detected. Thus the PFGE macrorestriction profiles reveal similarities and differences between isolates but provide little information on the genetic basis of isolate relationships. For example, Kudva *et al.* (2002) used PCR and Southern hybridizations to demonstrate that differing PFGE patterns in *E. coli* O157:H7 isolates resulted from discrete insertions or deletions in specific genomic regions characteristic of O157. In *Campylobacter*, PFGE pattern differences among clinical *C. jejuni* strains obtained from a waterborne Canadian outbreak were found to be the result of insertion or deletion events related to a Mu-like bacteriophage (Barton *et al.*, 2007).

CGI provided the highest level of discrimination between isolates and specific information on the genetic loci that accounted for the observed differences among them. Unique types were determined for all 36 isolates and as with PFGE there was a distinct tendency for the cattle and chicken isolates to cluster with isolates from the same host source. In addition, the genotypic relationships determined by PFGE were quite well conserved by CGI, including the 11 *C. jejuni* isolates with common PFGE types. Rodin *et al.* (2008) also reported a similar agreement between CGI and PFGE clustering results for *C. jejuni* isolates although clustering by host source was not conspicuous. Taken together, these results show that similar PFGE patterns are indicative of common genetic backgrounds that CGI can elucidate.

The *flaA* SVR method has proven to be a useful and convenient sequence based typing technique for *Campylobacter* (Meinersmann *et al.*, 1997, 2005; Dingle *et al.*, 2005) and is particularly useful for studies of recent outbreaks or closely related strains (Sails *et al.*, 2003). In our study, *flaA* SVR differentiated 35 unique types and one common type, appreciably better isolate discrimination than found with PFGE. However, the genetic relationships determined using *flaA* SVR shared less with those identified by CGI and PFGE, and only cattle isolates clustered by host source with *flaA* SVR. The high degree of variability inherent in the SVR of *flaA* thus proved useful for isolate discrimination but did not reflect the larger genomic scales of the CGI and PFGE techniques.

Eleven main categories for regions of high genetic variability were identified from the CGI data, comprising a total of 470 genes. Database searches of the annotated *C. jejuni* genomes NCTC11168 and RM1221 revealed that most of these regions are associated with established cellular functions including cell envelope components, transport and binding proteins, and energy metabolism. Also included were regions of mixed cellular functions and hypothetical proteins. Together with this group of variable genes were large insertion elements and recognized variable regions. Previous studies have identified four genomic islands (CJIE1–CJIE4) in RM1221 not found in NCTC11168 that have been confirmed by microarray analysis (Parker *et al.*, 2006) and sequence comparisons (Fouts *et al.*, 2005). These integrated elements have been studied in wild-type isolates and at least one of the regions was found in approximately 55% of *C. jejuni* isolates studied (Parker *et al.*, 2006). A genetic comparison of clinical *C. jejuni* isolates from South Africa also demonstrated a significant role for these unique elements in distinguishing closely related strains (Quiñones *et al.*, 2008). We found that integrated elements CJIE1, CJIE2, and CJIE4 represented a major portion (62%) of the variable genes attributed to RM1221. Smaller regions of variable genes in *C. jejuni* have also been shown to be important in defining differences among isolates (Pearson *et al.*, 2003; Taboada *et al.*, 2004; Parker *et al.*, 2006; Rodin *et al.*, 2008). The variable genes identified in our study included all or a segment of variable regions 1–17 (Parker *et al.*, 2006) and portions of recently identified regions 19 and 21 (Rodin *et al.*, 2008). Both the genomic islands and distinct variable regions would undoubtedly contribute significantly to a streamlined *Campylobacter* microarray design involving only key variable genes.

Multi-locus sequence typing (MLST) is a DNA sequence-based typing technique that evolved from multi-locus enzyme electrophoresis (Urwin and Maiden, 2003; Maiden 2006). A MLST

method involving sequence analysis of seven housekeeping genes was developed for *C. jejuni* (Dingle *et al.*, 2001, 2002), and MLST has also been used recently for the analysis of *C. coli* from human and food animal sources (Dingle *et al.*, 2005; Miller *et al.*, 2006). This technique has gained wide acceptance as a typing method for *Campylobacter*, owing in part to the establishment of a large *Campylobacter* MLST sequence type database (http://pubmlst.org). The housekeeping genes used in the MLST schemes are not hypervariable or subject to unusual selective forces and are therefore well suited to evolution and population studies of *Campylobacter* (Maiden, 2006). However, it is the reliance on a small number of relatively stable genes that limits MLST discrimination in some cases. For instance, in a study of *C. coli* from cattle, chickens, swine, and turkeys, 83% of the cattle isolates had the same MLST sequence type, ST-1068 (Miller *et al.*, 2006). A recent comparison study of genotypic and phenotypic methods for typing *Campylobacter* clinical isolates concluded that MLST was best suited for weakly clonal populations; supplementing MLST data with that from an additional genotypic method such as *flaA* SVR or ribotyping produced the best isolate discrimination (O'Reilly *et al.*, 2006). A similar conclusion was reached by Rodin *et al.* (2008) using oligonucleotide microarrays, PFGE, and MLST for genotyping *C. jejuni* from humans and chickens. They suggested combining MLST with a second typing method when examining epidemiological relationships of isolates over a limited time period. These observations point to the difficulty in designing a universal *Campylobacter* genotyping system equally useful for studying evolution as well as localized outbreaks.

While no bacterial typing method can answer every analytic need, CGI provides a wealth of genome-wide information for comprehensive analyses from which finer distinctions can be drawn. The cost of CGI analysis is comparable to other more technically involved genotyping methods such as PFGE (approximately US$40.00 per sample). However, whole genome analysis using DNA microarrays allows comparisons of the entire genetic content of different bacterial isolates, thus providing specific information on the genetic basis of the observed strain differences. Analysis of these differences can identify specific genes or gene clusters to serve as the basis for further, more detailed analysis of strain differences. As the number of sequenced *Campylobacter* genomes increases, focusing on these variable regions will enable the design of even more informative, yet simplified and more cost-effective *Campylobacter* microarrays with substantially reduced numbers of features. The inherent flexibility of the microarray format facilitates designs that can include multiple *Campylobacter* strains and species to accommodate diverse typing tasks. In the future more economical microarray platforms, labeling and hybridization technologies will further improve the cost effectiveness of DNA microarrays as tools for research and clinical investigations.

Acknowledgments

The authors thank Sandra House, Leena Jain, and Jodie Plumblee of the Bacterial Epidemiology and Antimicrobial Resistance Research Unit for excellent technical assistance. This work was supported by the U.S. Department of Agriculture-Agricultural Research Service CRIS projects 6612-32000-035 and 5325-42000-045. The mention of trade names or commercial products in this manuscript is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture.

Disclosure Statement

No competing financial interests exist.

References

Anjum MF, Lucchini S, Thompson A, *et al.* Comparative genomic indexing reveals the phylogenomics of *Escherichia coli* pathogens. Infect Immun 2003;71:4674–4683.

Barton C, Ng LK, Tyler SD, *et al.* Temperate bacteriophages affect pulsed-field gel electrophoresis patterns of *Campylobacter jejuni*. J Clin Microbiol 2007;45:386–391.

Brown PO and Bostein D. Exploring the new world of the genome with DNA microarrays. Nat Genet 1992;21:25–32.

Call DR, Borucki MK, and Besser TE. Mixed-genome microarrays reveal multiple serotype and lineage-specific differences among strains of *Listeria monocytogenes*. J Clin Microbiol 2003;41:632–639.

Dingle KE, Colles FM, Falbush D, *et al.* Sequence typing and comparison of population biology of *Campylobacter coli* and *Campylobacter jejuni*. J Clin Microbiol 2005;43: 340–347.

Dingle KE, Colles FM, Ure R, *et al.* Molecular characterization of *Campylobacter jejuni* clones: a basis for epidemiologic investigation. Emerg Infect Dis 2002;8: 949–955.

Dingle KE, Colles FM, Wareing DR, *et al.* Multilocus sequence typing system for *Campylobacter jejuni*. J Clin Microbiol 2001;39:14–23.

Ehrenreich A. DNA microarray technology for the microbiologist: an overview. Appl Microbiol Biotechnol 2006;73:255–273.

Englen MD and Fedorka-Cray PJ. Evaluation of a commercial diagnostic PCR for the identification of *Campylobacter jejuni* and *Campylobacter coli*. Lett Appl Microbiol 2002;35:353–356.

Fouts DE, Mongodin EF, Mandrell RE, *et al.* Major structural differences and novel potential virulence mechanisms from the genomes of multiple campylobacter species. PLoS Biol 2005;1:e15.

Friedman CR, Neimann J, Wegner HC, *et al.* Epidemiology of *Campylobacter jejuni* infections in the United States and other industrialized nations. In: Campylobacter, 2nd edition. Nachamkin I and Blaser MJ (eds). Washington, DC: American Society for Microbiology, 2000, pp. 121–138.

Garaizar J, Rementeria A, and Porwollik S. DNA microarray technology: a new tool for the epidemiological typing of bacterial pathogens? FEMS Immunol Med Microbiol 2006;47:178–189.

Gibson J, Lorenz E, and Owen RJ. Lineages within *Campylobacter jejuni* defined by numerical analysis of pulsed-field gel electrophoretic DNA profiles. J Med Microbiol 1997;46:157–163.

Hiett KL, Stern NJ, Fedorka-Cray PJ, *et al.* Molecular phylogeny of the *flaA* short variable region among *Campylobacter jejuni* isolates collected during an annual evaluation of poultry flocks in the southeastern United States. Foodborne Pathog Dis 2007;4:339–347.

Kostrzynska M and Bachand A. Application of DNA microarray technology for detection, identification, and characterization of food-borne pathogens. Can J Microbiol 2006;52:1–8.

Kudva IT, Evans PS, Perna NT, *et al.* Strains of *Escherichia coli* O157:H7 differ primarily by insertions or deletions, not single nucleotide polymorphisms. J Bacteriol 2002; 184:1873–1879.

Leonard II EA, Takata T, Blaser MJ, *et al.* Use of an open-reading frame-specific *Campylobacter jejuni* DNA microarray as a new genotyping tool for studying epidemiologically related isolates. J Infect Dis 2003;187: 691–694.

Maiden MCJ. Multilocus sequence typing of bacteria. Annu Rev Microbiol 2006;60:561–588.

Mead PS, Slutsker L, Dietz V, *et al.* Food-related illness and death in the United States. Emerg Infect Dis 1999;5:607–625.

Meinersmann RJ, Helsel LO, Fields PI, *et al.* Discrimination of *Campylobacter jejuni* isolates by *fla* gene sequencing. J Clin Microbiol 1997;35:2810–2814.

Meinersmann RJ, Phillips RW, Hiett KL, *et al.* Differentiation of *Campylobacter* populations as demonstrated by flagellin short variable region sequences. Appl Environ Microbiol 2005;71:6368–6374.

Miller WG, Englen MD, Kathariou S, *et al.* Identification of host-associated alleles by multilocus sequence typing of *Campylobacter coli* strains from food animals. Microbiology 2006;152:245–255.

Moore JE, Corcoran D, Dooley JS, *et al. Campylobacter*. Vet Res 2005;36:351–382.

Ochman H and Santos SR. Exploring microbial microevolution with microarrays. Infect Genet Evol 2005; 5:103–108.

O'Reilly LC, Inglis TJJ, and Unicomb L. Australian multicentre comparison of subtyping methods for the investigation of *Campylobacter* infection. Epidemiol Infect 2006;134:768–779.

Parker CT, Quiñones B, Miller WG, *et al.* Comparative genomic analysis of *Campylobacter jejuni* strains reveals diversity due to genomic elements similar to those present in *C. jejuni* strain RM1221. J Clin Microbiol 2006; 44:4125–4135.

Parkhill J, Wren BW, Mungall K, *et al.* The genome sequence of the food-borne pathogen *Campylobacter jejuni* reveals hypervariable sequences. Nature 2000;403:665–668.

Patton CM, Wachsmuth IK, Evins GM, *et al.* Evaluation of 10 methods to distinguish epidemic-associated *Campylobacter* strains. J Clin Microbiol 1991;29:680–688.

Pearson BM, Pin C, Wright J, *et al.* Comparative genome analysis of *Campylobacter jejuni* using whole genome DNA microarrays. FEBS Lett 2003;554:224–230.

Quiñones B, Guilhabert MR, Miller WG, *et al.* Comparative genomic analysis of clinical strains of *Campylobacter jejuni* from South Africa. PLoS ONE 2008;3:e2015.

Ribot EM, Fitzgerald C, Kubota K, *et al.* Rapid pulsed-field gel electrophoresis protocol for subtyping of *Campylobacter jejuni*. J Clin Microbiol 2001;39:1889–1894.

Rodin S, Andersson AF, Wirta V, *et al.* Performance of a 70-mer oligonucleotide microarray for genotyping of *Campylobacter jejuni*. BMC Microbiol 2008;8:73.

Sails AD, Swaminathan B, and Fields PI. Utility of multilocus sequence typing as an epidemiological tool for investigation of outbreaks of gastroenteritis caused by *Campylobacter jejuni*. J Clin Microbiol 2003;41:4733–4739.

Singer RS, Sischo WM, and Carpenter TE. Exploration of biases that affect the interpretation of restriction fragment patterns produced by pulsed-field gel electrophoresis. J Clin Microbiol 2004;42:5502–5511.

Skirrow MB. Diseases due to *Campylobacter*, *Helicobacter* and related bacteria. J Comp Pathol 1994;111:113–149.

Stern NJ, Wojton B, and Kwiatek K. A differential-selective medium and dry ice-generated atmosphere for recovery of *Campylobacter jejuni*. J Food Prot 1992;55:514–517.

Taboada EN, Acedillo RR, Carrillo CD, *et al.* Large-scale comparative genomics meta-analysis of *Campylobacter jejuni* isolates reveals low level of genome plasticity. J Clin Microbiol 2004;42:4566–4576.

Tenover FC, Arbeit RD, Goering RV, *et al.* Interpreting chromosomal DNA restriction patterns produced by pulsed-field gel electrophoresis: criteria for bacterial strain typing. J Clin Microbiol 1995;33:2233–2239.

Urwin R and Maiden MCJ. Multi-locus sequence typing: a tool for global epidemiology. Trends Microbiol 2003; 11:479–487.

Wassenaar TM and Newell DG. Genotyping of *Campylobacter* spp. Appl Environ Microbiol 2000;66:1–9.

Address reprint requests to:
Mark D. Englen, M.Sc., Ph.D.
Bacterial Epidemiology and Antimicrobial Resistance Research Unit
Agricultural Research Service
U.S. Department of Agriculture
Richard B. Russell Agricultural Research Center
950 College Station Rd.
Athens, GA 30605-2720

E-mail: mark.englen@ars.usda.gov